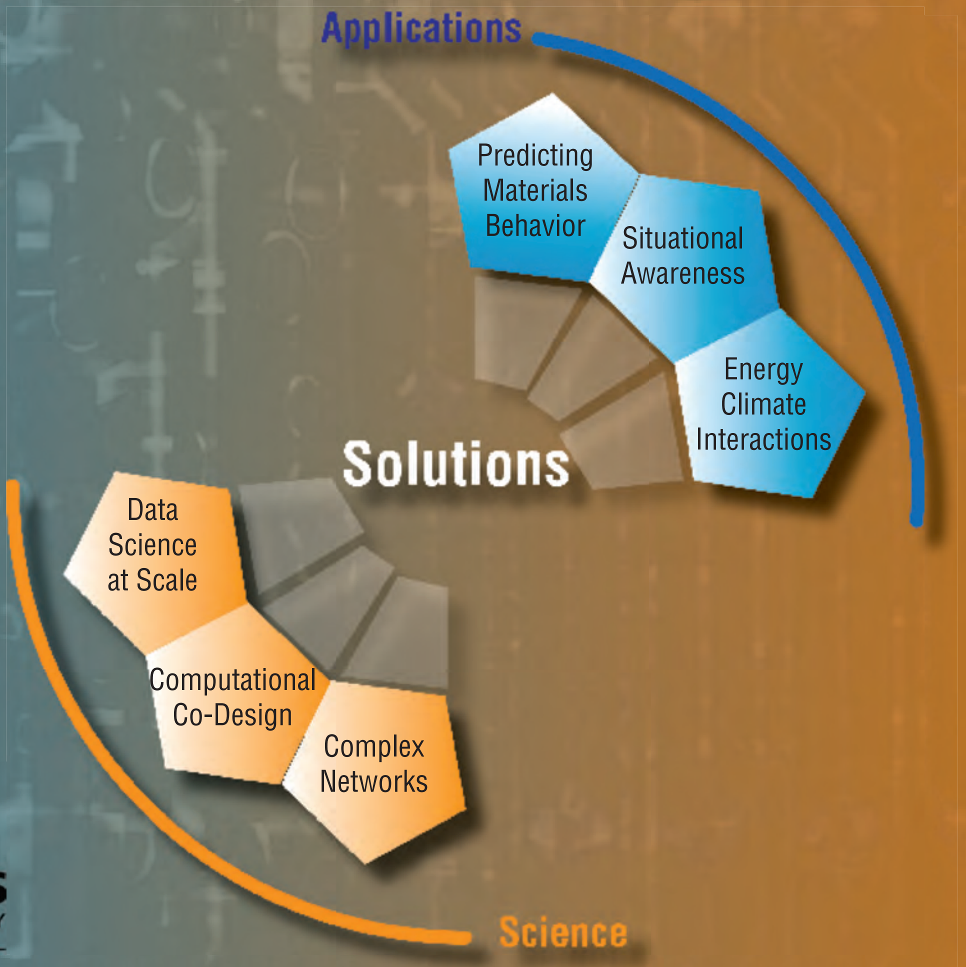
Science That Matters:

Integrating Information, Science, and Technology for Prediction

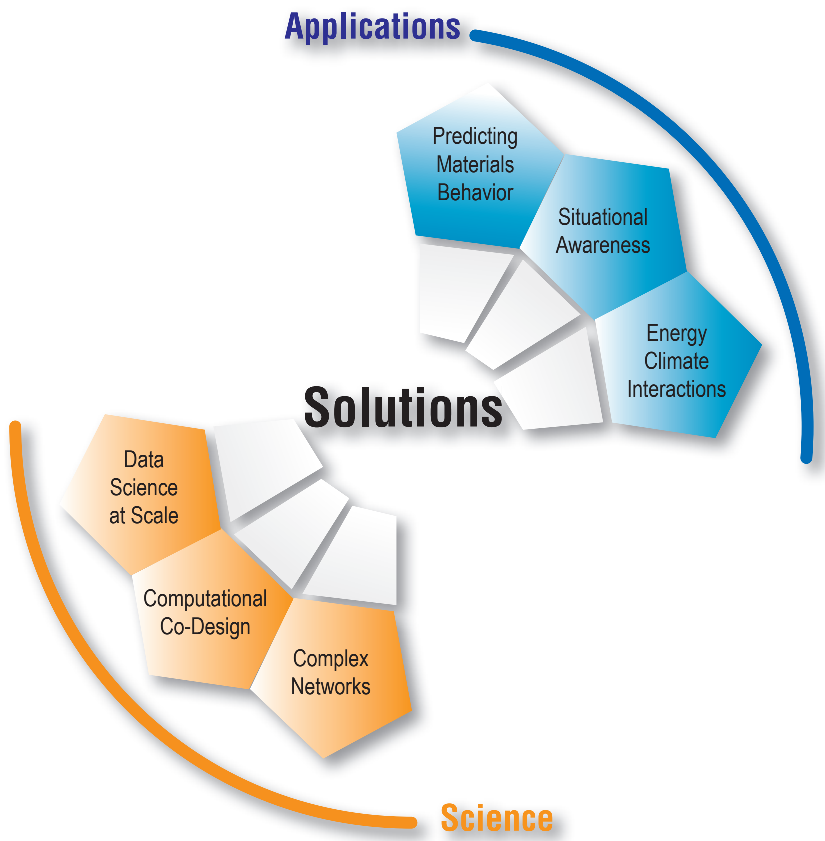
Strategic Plan

December 2010



Los Alamos NATIONAL LABORATORY EST. 1943

LA-UR 11-00115



LANL's capability pillar "Integrating Information, Science, and Technology for Prediction" addresses emerging challenges in national security, societal prosperity, and fundamental science. This pillar leverages advances in theory, algorithms, and the exponential growth of high-performance computing to accelerate the integrative and predictive capability of the scientific method. The Information Science and Technology (IS&T) thrust will focus on the integration of LANL assets for understanding, quantified prediction, and design of complex natural and engineered systems.

This document was coordinated by the IS&T Center and contains input from technical staff from across the Laboratory. See http://institutes.lanl.gov/istc/board-of-directors/ *for the current Board of Directors and Science Council.*

Contents

LA-UR 11-00115

Introduction

The new millennium presents us with a number of emerging challenges such as energy security, climate change, international terrorism, the aging of our nuclear deterrent, and the proliferation of nuclear weapons. These global challenges are notable for their daunting scale as well as for their extreme level of complexity. Our significant national assets in observational and experimental data acquisition, computational capacity, and modeling and simulation are already addressing these challenges; however, the emerging application frontiers mentioned above are revolutionary in their requirements.

As a result, we need to exercise the scientific method at a scale of complexity that is only now becoming feasible. To meet these emerging global challenges, we must be deliberate in our investment of resources and intellectual capital to integrate interdisciplinary skills with the goals of understanding and of enhancing tools for prediction and design. Such integration has been a hallmark of the Los Alamos National Laboratory throughout its history. Due to the nature and scale of our core weapons design and stewardship mission, we have learned to emphasize interdisciplinary approaches and often thereby changed the nature and frontiers of certain disciplines. This century's national challenges now call us to take this tradition to even higher levels and lead the world in integrative Science, Technology, and Engineering (ST&E).

Achieving such leadership requires the integration of discovery science, validated theory, computational algorithms, software infrastructure, and computing hardware. This integrated "toolset" must be particularly suited to the advanced computing architectures of the future, which will likely be increasingly heterogeneous and multicore and will be designed for both compute-intensive and data-intensive applications. Success demands a self-consistently designed system of theory, computation, and experiment and observation. Achieving such "co-design" at scale is central to addressing the compelling national security issues with the generation of new ideas, concepts, and methodologies to enhance the fidelity, reliability, certainty, and usability of codes – to guide and interpret experiments, and provide to prediction and design control for complex phenomena and systems.

Integration, for the purpose of discovery and prediction, requires the full coupling of advanced modeling and simulation codes and Information Science and Technology (IS&T) infrastructure with data from many sources, including suites of experiments and observations. Here data plays the traditional role of motivating theories and validating model predictions, which in turn play roles of interpreting and motivating experiments. The pillar "Integrating Information, Science and Technology for Prediction" aims to pave the way to an even more integrated co-design strategy for a discovery-to-prediction Laboratory.

The timing of this opportunity and challenge is fortuitous – we are entering a period of disruptive change in high-performance technology, several major national experimental and data facilities are being contemplated, and the country has invested high expectations in ST&E to provide solutions to the pressing national challenges. Capabilities in data intensive science, computational co-design, and complex networks can greatly enhance our impact on solving the most challenging problems facing our nation.

Vision

Our goal is to establish LANL as the IS&T leader within the Department of Energy (DOE), and as a national resource for national security science. The LANL IS&T capability will integrate all aspects of experimental, analytical, and computational capabilities in support of scientists and decision makers.

The Challenge of Integrating Information, Science, and Technology for Prediction

In the next decade sensors and experiments will be generating more data than ever before. For example, the Large Synoptic Survey Telescope (LSST) produces data at rates of up to 30 terabytes per night. At the Large Hadron Collider a number of experiments are expected to run for many years with data rates of up to 1.5 GB per second, yielding up to 100 petabytes of data. DOE genomics programs, such as Genomes-To-Life, will analyze hundreds of petabytes of data per year. Defense and intelligence data sources may dwarf even these scientific applications and have the additional challenge of a real-time analysis component. Additionally, digital audio, video, and image data volumes are growing rapidly both in scientific applications and in the commercial sector. The IS&T challenges associated with the management and analysis of this explosion of data are daunting, but must be surmounted in order for LANL to fulfill its scientific and national security missions.

Los Alamos Contributions to Integrating Information, Science, and Technology for Prediction

Los Alamos National Laboratory's IS&T capability builds upon our decades-long expertise and rich tradition in integrating high performance computing, theory, statistics, learning, modeling, simulation, and the instrumentation and diagnostics of scientific experiments. This IS&T history includes the development of the Monte Carlo method, originating work in the Human Genome Project, pioneering contributions to nonlinear dynamical systems, and network theory and applications. More recently, LANL has been at the forefront in the fast-growing field of quantum IS&T. We cannot rest on these and other past accomplishments, however. IS&T is rapidly changing, and we must not only keep pace with external advances, but take a leadership position in the specific areas necessary to ensure LANL's future successes.

LANL's thrust in IS&T is developing new mathematical and computational approaches to apply advanced information sciences and technologies to problems of national importance. The thrust will focus on advancing our IS&T capabilities in support of the DOE and national missions by catalyzing the DOE workforce through extensive collaborations between national laboratories, academic institutions, government agencies, and industries with complementary expertise. In addition, we will accelerate communication between IS&T advances over a wide array of application domains. These advances are leveraged by the underlying commonality of IS&T needs and applicability to our multiple missions.

The DOE is a world leader in the analysis of large-scale data, using high-end computer simulations to create accurate and predictive models of physical systems. The quality of our science-based discovery and predictions rests on effective IS&T tools to connect the three pillars of the scientific paradigm – theory, experiment, and simulation.

The success of the LANL IS&T capability will directly impact Laboratory accomplishments and create new opportunities for growth. The metrics for success will be measured by its direct impacts on multiple DOE missions, and in particular, on Laboratory missions for ensuring the safety, security, and reliability of the US nuclear deterrent, reducing global threat, and solving emerging national security challenges, including coupled energy-climate-infrastructure management.

Our national security and economic competitiveness require that the United States remain at the forefront of IS&T. Therefore, we must be cognizant of the international advances and opportunities and must be able to anticipate surprises in technology. We at LANL we must be able to collaborate effectively with the world community in a multinational, open scientific environment. LANL has an exceptional track record of applying basic, advanced unclassified research to classified applications. The ability to support open, cutting-edge, multinational science and apply these advances to classified national security problems will be key to our continued leadership.

LANL has the infrastructure and experienced workforce necessary to address the challenges of IS&T in the short term, and to develop the capabilities and tools needed in the long term. LANL's unique IS&T environment will enable science that can address a broad set of needs. We will create a unique and compelling capability, and we will set the standard by which IS&T efforts are measured. LANL will considerably extend the state of the art, so that its facilities and expertise will be highly sought by scientists worldwide.

The LANL capability pillar Integrating Information, Science, and Technology for Prediction will directly support DOE's mission to deliver capabilities to scientists nationwide that enable them to extend the frontiers of science, and will provide the ongoing support for the US nuclear deterrent required by the Nuclear Posture Review (NPR).

Strategy

The LANL pillar, Integrating Information Science and Technology for Prediction, will be directed at solving important national problems. The following are the three principal elements to this pillar:

Identification of Critical National Need – This element requires understanding those areas of critical need where LANL can make significant contributions that are synergistic with our capabilities and other national security missions.

Maintenance and Enhancement of Capabilities – LANL must sustain its science and technology capabilities base, including both human capital and facilities.

Development of New Programs – Our success will be measured by our ability to grow a significant programmatic base through transformational ideas in critical areas where LANL has a comparative advantage. To build new programs, we will assure high-quality execution on our current programs and engage DOE and other sponsors at multiple levels.

We have identified the following three IS&T application areas where LANL's unique capabilities in IS&T will allow us to make important and distinguishing contributions:

- Predicting Materials Performance
- Situational Awareness
- Energy-Climate Impacts and Energy Infrastructure

To address these application areas, LANL requires strong capabilities in the following three common, cross-cutting IS&T areas:

- Data Science at Scale
- Computational Co-design
- Complex Network Science

The Los Alamos National Laboratory is committed to developing and strengthening these underlying capabilities, which are necessary to succeed in the three application areas listed above, as well as emerging challenges. Two overarching themes across capabilities exist. The first is quantification of uncertainty in our predictions—we must be able to speak to the certainty of our predictions.

The second overarching theme is that these predictions will often be used to support decision making in situations for which the predictions cannot be directly tested. Examples of such situations, for which direct experimental access is not available, include future forecasts of climate or biological response under varied scenarios, and predictions of nuclear weapon performance in the absence of nuclear testing.

The ability to predict in experimentally unreachable regimes places an enormous, almost transformational, burden on the scientific method. To employ the scientific method in such situations, we must restrict ourselves to theories and models that can be expected to extrapolate from the regimes that can be tested to the regime where a prediction is required. Such models tend to be more foundational and less empirical in approach. One must also develop computational infrastructure that can correctly simulate the interactions of such models at the scales required for prediction. In support of the weapons program, LANL has a long history of applying such techniques.

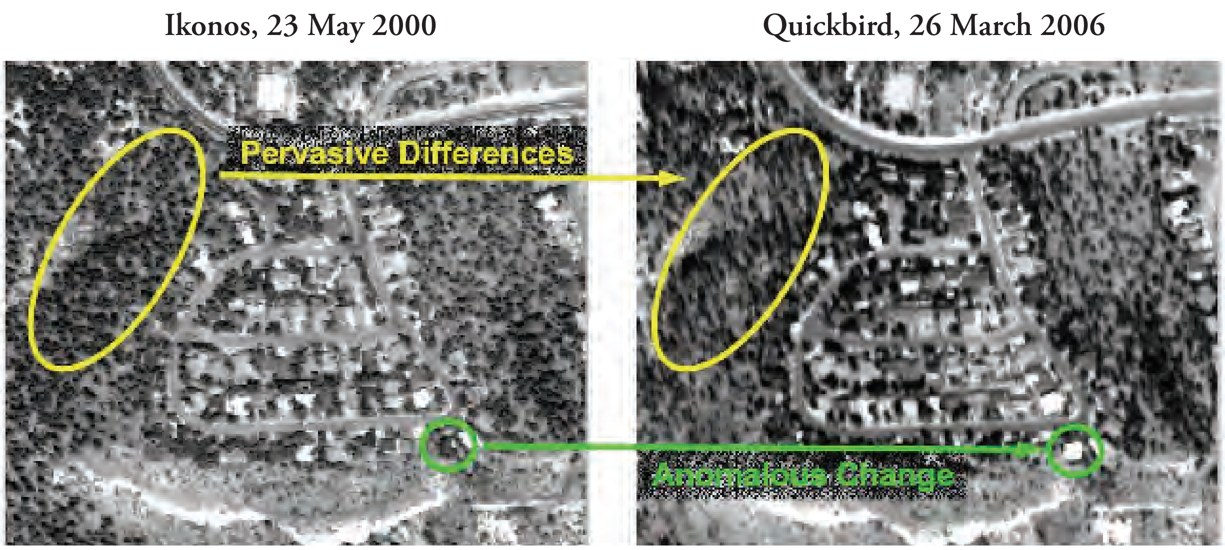
This plan is built upon a set of strategic goals that support the three research and development (R&D) application areas and underpin the three IS&T capabilities. Each of these application areas is scientifically challenging, as well as vitally important for national security. In addition, there exists a funding base for each area outside LANL, toward which we are targeting program growth. For each of these key areas, we have outlined a specific set of 5-to-10-year R&D goals. These goals are the result of a careful evaluation of where national needs and priorities intersect with our existing and growing capabilities in order to define how LANL can make the greatest contribution to national IS&T. The implementation plan, found in a separate document, presents key activities over the next 1 to 3 years for achieving these long-term goals.

Scientific and Technical Challenges and LANL IS&T Goals

Capabilities

Success in each strategic application area requires the surmounting of underlying scientific and technological challenges. We have identified three overarching scientific areas or capabilities that need to be healthy in order for LANL to achieve its strategic goals: Data Science at Scale, Computational Co-design, and Complex Network Science. These capabilities are discussed in detail in the following sections.

A strong capability in Quantum Information Science (QIS) has been one of the goals of the IS&T pillar over the past few years. With significant internal investment and development, the QIS capability has matured and is now entering a new and transitional and



The distinction between pervasive differences and anomalous changes.

translational phase. QIS is widely recognized as a disruptive technology with the possibility to rapidly change the technology landscape in areas that will continue to be important to the Laboratory, such as information security, simulation and design of materials, and state-of-the-art sensing. As such it will continue to be nurtured and guided. QIS efforts will be coordinated by the IS&T Center and the Center for Non-linear Science (CNLS).

In a similar fashion, annual updates to the strategic plan will reflect transitions of the above three overarching capabilities to new phases in their development cycles.

Capability: Data Science at Scale

Extremely large datasets and extremely high-rate data streams are becoming increasingly common due to the operation of Moore's Law as applied to sensors, embedded computing, and traditional high-performance computing. Interactive analysis of these datasets is widely recognized as a new frontier at the interface of information science, mathematics, computer science, and computer engineering. Text searching on the web is an obvious example of a large dataset analysis problem; however, scientific and national security applications require far more sophisticated interactions with data than text searches. These applications represent the "data to knowledge" challenge posed by extreme-scale datasets in, for example, astrophysics, biology, climate modeling, cyber security, earth sciences, energy security, materials science, nuclear and particle physics, smart networks, and situational awareness.

In order to contribute effectively to LANL's overall national security mission, we need a strong capability in Data Science at Scale. This capability rests on robust and integrated efforts in data management and infrastructure, visualization and analysis, high-performance computational statistics, machine learning, uncertainty quantification, and information exploitation. The Data Science at Scale capability provides tools capable of making quantifiably accurate predictions for complex problems with the efficient use and collection of data and computing resources.

Data Management and Infrastructure

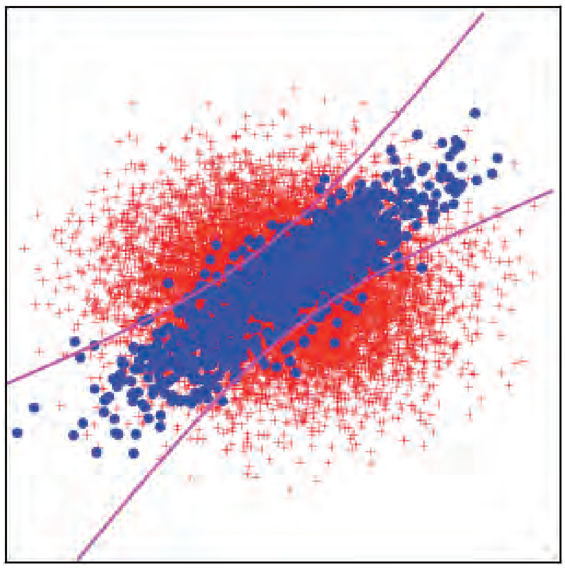
In order to effectively process the massive, mission-created datasets that will be generated in the near future, a robust effort that researches, designs, and implements data-handling software and hardware architectures is required. To meet this challenge, LANL will continue to enhance its capabilities in this foundational area.

Visualization and Analysis

The human visual system is one the most effective ways to convey information. Visualization methods transform datasets from raw numbers to geometric and image-based representations that help people understand their data. LANL's visualization tools are used by scientists worldwide to visualize their massive datasets, and LANL will maintain and advance its already strong program in this area.

Statistical Sciences

The current LANL portfolio of statistical work includes a focus on nuclear weapons, with significant efforts in threat reduction, energy, and science. Within this mission space, state-of-the-art statistics is incorporated into collaborative efforts that address several of the goals set out in the Laboratory's grand



The hyperbolic change detector. The horizontal axis corresponds to one image and the vertical axis to the other. The pervasive differences (blue dots) are obtained from corresponding pixels in the two images; the anomalous changes (red crosses) are modeled to exhibit the same distribution over the individual images but to be uncorrelated across the images. For Gaussian data, the magenta lines that optimally separate pervasive differences from anomalous changes are hyperbolae.

challenges areas including (but not limited to) data integration, multiscale modeling, infrastructure modeling, nonproliferation, counterterrorism, social modeling, material attribution, cosmology, weapons physics evaluation needs, cyber security, and biological threats. To deal with massive amounts of static and streaming data, LANL is strengthening its high performance computational statistics capability.

Machine Learning

Machine learning encompasses nearly all types of inference, but LANL researchers have focused specifically on detection and classification problems. Detection and classification problems are ubiquitous – for example, we would like to detect and classify threats, disease, fraud, anomalies, proliferation, intrusions, military targets, improvised explosive devices (IEDs), biomarkers, etc. Although the specifics of detection and classification can be quite diverse, there are common, fundamental issues that must be overcome for success. These issues include limited first principles information, limited empirical information (i.e., ground truth data) and limited computational resources. Building a strong program in machine-learning-based detection science to address this class of problems will have significant impact on national security and the human condition, and will play a critical role in enabling scientific discovery.

Uncertainty Quantification

Uncertainty Quantification (UQ) consists of the suite of tools used to synthesize the output of (typically large-scale) computation with experimental and observational data for the purpose of drawing inferences and making predictions. In addition to these inferences and predictions, UQ provides an associated measure of confidence. LANL has a rich history of applying UQ to diverse areas such as weapons physics, cosmology, energy, and hydrology, and we are currently extending this set to include environmental management and climate. In the coming years we will further extend the applicability of this framework to problems beyond the physical systems to which UQ was originally applied. These new application domains include, among others, cyber security, remote sensing, materials science, and epidemic modeling. UQ must be an integral part of all of our modeling and simulation work at LANL.

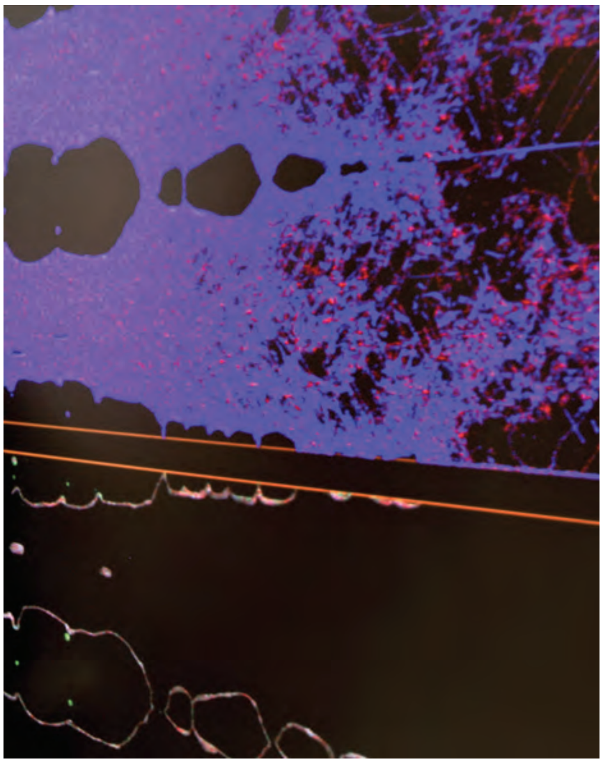
To achieve a strong Data Science at Scale program, an integrated approach that draws from these areas (data management, visualization, etc.) is required. A vibrant program will develop from an interdisciplinary approach that overcomes traditional silos in which each of these areas is a separate, independent task. For example, a key capability for the Laboratory in the future is information exploitation. Here human experts use a tremendous amount of prior information and contextual cues to interpret complex data. Exactly how to build automated systems with these capabilities, however, is a long-standing problem in information science. Solutions to this problem would provide significant improvements in human productivity, vast reduction in data storage and exploitation costs, and an ability to digest and analyze orders of magnitude more data than the current capability. Successful solutions to this problem will necessarily require progress on multiple fronts, including data management and infrastructure, visualization, inference and learning, knowledge management, decision support, and human-computer interaction. These solutions would support collaborative and distributed data exploitation, and be scalable in both the volume and diversity of data. LANL's significant expertise and resources in high performance computing, coupled with our expertise in statistical and semantic analysis, our decision support systems capability, our library science, and our considerable and often unique subject matter expertise will be leveraged to solve outstanding problems in information exploitation.

Goals for Data Science at Scale are as follows:

- Position Los Alamos National Laboratory to compete for external program opportunities involving massive datasets and streaming data by:
  - fostering the scientific underpinnings of data intensive computing, including streaming data
  - launching a Data Science at Scale laboratory for hardware/software/algorithmic experimentation with massive datasets
  - building scalable data processing capabilities in statistics, UQ, and machine learning
  - collaborating with strategic external partners in the area of data science
- Position LANL to compete for external program opportunities requiring information exploitation by:
  - establishing practical yet rigorous detection science as part of a targeted global security program
  - developing a nationally recognized program and becoming the lead national laboratory in inference and learning and UQ
  - building integrated efforts that jointly solve problems in the area of information exploitation (for example, visualization/machine learning, statistics and data infrastructure)

Capability: Computational Co-design

Computational Co-design is the process of designing the interacting components of a computational system as a whole, rather than separately optimizing each component based on



SPaSM, Scalable Parallel Short-range Molecular-dynamics code, ported to Roadrunner supercomputer static requirements and interface descriptions from the other components. These components span the spectrum from low-level hardware to theory, and include system software, programming models, programming languages, algorithms, and even data analysis and visualization. The expectation is that enlarging the design space will allow for significantly better, perhaps even revolutionary, design. The co-design process can be very complex. It is a multidimensional, constrained optimization problem, the solution of which is demanding in its own right.

Co-design is by no means new and has been formally developed for the design of embedded information systems in automobiles, aircraft, and other applications. At LANL we are interested in the co-design of large-scale complex systems. In a broader context, co-design is central to the DOE exascale initiative currently targeted to begin in fiscal year 2012. Through co-design we will bind together applications, methods, algorithms, and the supporting software and hardware to achieve at least a 1000-fold increase in computational capability. Broader still, this will be a decade of opportunity—revolutionary experiments are planned in areas such as materials, climate, cosmology, high-energy physics, and fusion energy. Predictive modeling and simulation promises to provide dramatic new capabilities for both understanding and decision-making – each is dependent on the other.

Often in the past, computing and information technology decisions have been made based on static requirements – such as bandwidth, memory size, and processor speed. Over the next decade, the increased ability to compute, analyze, store, and communicate data will transform the traditional scientific and engineering enterprise. This rapid change will generate new challenges – reducing the power consumption of supercomputers and the research and development of new algorithms and techniques – as well as new opportunities – quantifying the uncertainty of complex simulations and providing reliable predictions in experimentally inaccessible regimes.

At LANL we will implement the co-design paradigm for applications in materials performance, energy-climate systems analysis, and situational awareness problems. We will use this paradigm to broaden the domain of reliability of predictions of nuclear weapon performance, thus supporting the NPR conclusion that investment was required for: "Strengthening the science, technology, and engineering (ST&E) base needed for conducting weapon system life extension plans, maturing technologies to increase weapons surety, qualification of weapon components and certifying weapons without nuclear testing, and providing stockpile assessments through weapons surveillance. This includes developing and sustaining high quality scientific staff and supporting computational and experimental facilities."

Steps along this path at LANL include the following:

- Co-designing extreme computing architectures, algorithms, and methods tailored to the classes of new problems posed by cutting-edge applications for current and future national security missions and facilities. The new decadal DOE-wide initiative for exascale computing will fully exercise this level of integration.
- Making computer and computational science integral elements of major new experimental facilities and programs. Current examples in planning phases at LANL include MaRIE (Matter-Radiation Interactions in Extremes), ECI (Energy and Climate Impacts), smart grids, situational awareness, cyber security, intelligence, and biology.
- Using modern information and visualization technology to enhance active data management and exchange within extensive synthesis/data/characterization/theory, modeling, and simulation teams.

- Using tools from our Data Science at Scale thrust to extract key features from massive datasets, inform optimal models for analysis and simulation, and incorporate UQ/ Verification & Validation (UG/V&V) methodologies to suggest the most effective experiments for evaluation of a given theoretical framework.
- Facilitating synergy and exchange of IS&T and code capabilities and strategies between programs, and thereby accelerating agile, innovative responses to national security challenges.

Goals for Computational Co-design are as follows:

- Establish LANL's leadership role in computational co-design.
- Ensure that co-design is an integral part of the MaRIE signature facility.
- Develop a critical mass of technical staff with co-design expertise.
- Secure a broad portfolio of externally funded programs in computational co-design.
- Develop a generalized process of co-design to bind together experiments, applications, and technology and create a truly integrated science and technology enterprise at LANL

Capability: Complex Network Science

Networks provide a natural framework for describing complex systems in terms of interdependent subsystems and components. The network can be dynamic to reflect changes of either the components and/or the dependencies between them over time. Complex systems and their network representations arise in many scientific and national security mission areas at LANL: cyber systems, national infrastructure, biological systems, real and artificial neural systems, classical and quantum communication systems, systems of networked sensors, social networks, terrorist networks, and energy systems such as the smart grid.

Across this spectrum of complex networks there are common R&D frontiers that include discrete mathematics and graph-based methods, as well as specific algorithmic and computing architectures. It is by both advancing the state of the art in these relevant disciplines and by integrating them with one another, focused toward application in Complex Network Science, that we will advance this capability of the IS&T pillar with maximum impact.

To enhance LANL's capability in Complex Network Science, we will focus on the following areas:

- **Network Design** – Many networks arise through multiobjective optimization. This is particularly the case for engineered systems such as the power grid, cyber systems, and biological systems that have evolved to optimize both throughput and resilience. Developing mathematical and computational tools to perform such optimization is crucial and challenging because most of the optimization problems are computationally hard. As a result, we need to develop both accurate heuristics and probabilistic methods for optimization on networks.
- **Network Inference** – Many networks are too large to be observed all at once, or are changing with time. Examples of network inference problems are the identification of missing links in a graph, the discovery of the initial case of an epidemic, and the construction of a phylogenetic network from raw DNA sequences. Given the data, we need to be able to infer the generating mechanism of the network. Here, data refers to measured attributes for both nodes and links. As a result, the amount of data can grow as the square of the number of nodes, making this yet another example of data analysis at scale.
- **Network Modeling** – Both network design and inference rely on models for network evolution. Significant effort is needed to develop validated models for specific applications, and general computational tools

to generate networks from the models. With this modeling effort, we have the opportunity to efficiently transfer what we learn in one application area to other application areas, thereby realizing the promise of generality of an IS&T effort.

- **Network Dynamics** – In many examples, the network serves as a substrate on which a dynamical process of interest lives. One example is a communicable disease (or a computer virus) propagating on a social network. To understand, predict, and control such systems requires tools across fields – game theory, stochastic processes, statistical physics, graph theory, etc. Moreover, domain-specific expertise (e.g., biochemistry, economics, and computer science) is also required. Over the last decade, LANL's significant capability in statistical physics has been drawn upon to study such these systems, but as the systems become more complex we need to draw upon and develop expertise in the other areas of mathematics and computer science.

To meet these emerging challenges, a working group is examining the intersections between these and other significant disciplines within the applicable program mission areas, and will create a program plan charting a course for Complex Network Science in support of the LANL capability pillar. To achieve these goals our strategy will be to leverage our significant LDRD (Laboratory Directed Research and Development) investment and work with program offices to develop large-scale funded projects.

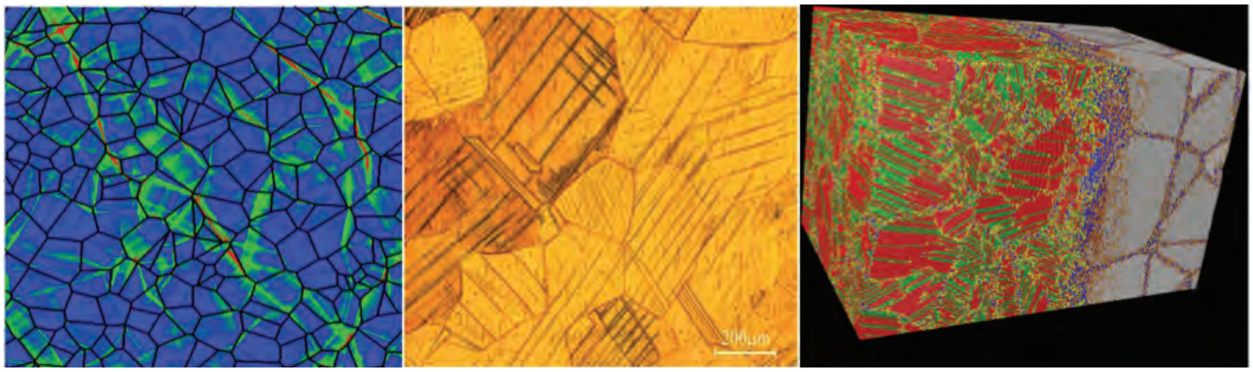
Goals for Complex Network Science are as follows:

- Establish LANL's leadership role across a spectrum of application areas in complex networks science including cyber security, smart-grid systems, and metagenomic systems.
- Establish LANL's leadership role in the DOE and government for cognitive neuroscience for national security, and position LANL to compete for external program opportunities requiring an advanced synthetic cognition capability.
- Achieve visual synthetic cognition with an artificial neural network, as measured by the ability to outperform a human in certain recognition tasks.
- Demonstrate quantum-enabled secure communication on a networked system of users.

Application Areas

Application Area: Predicting Materials Performance

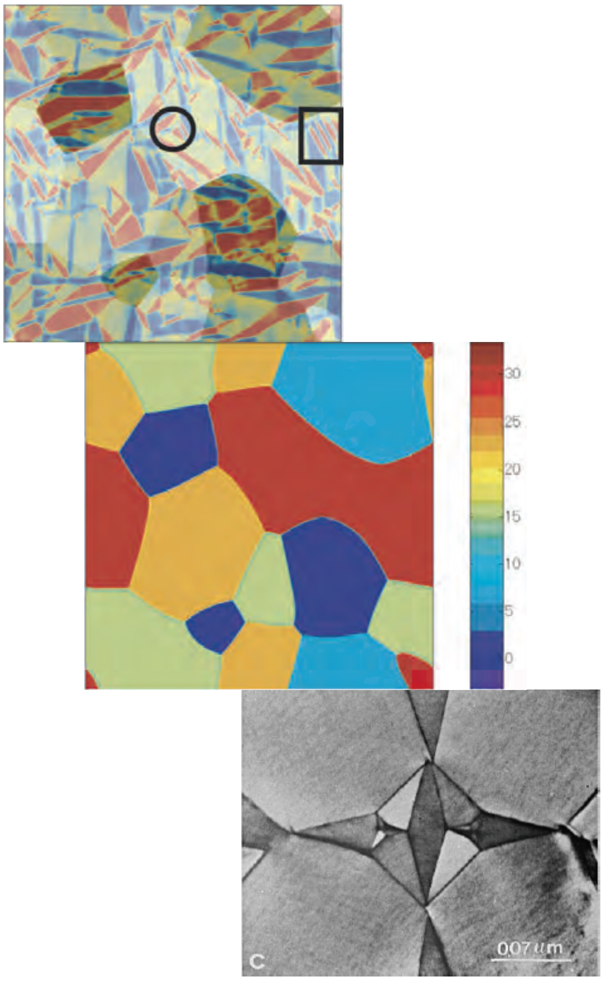
The development of a predictive capability for materials performance, coupled with a realistic assessment of model and prediction uncertainties, is one of the key challenges underlying LANL's mission. The understanding of a material's heterogeneities and fluctuations at a wide variety of length and time scales and their influence on constitutive response are key to addressing this challenge. The heterogeneities can range from atomistic defects (<1 nm) through coupled defect structures (dislocation loops, twin planes ~100 nm) up through the polycrystalline morphology (>1 micron) of the materials. The initial distributions of such structures are very much process dependent, though our capabilities to design and attain a particular configuration are limited and empirically based. Advancement in this area will require a well-validated, integrated, multiscale modeling approach. Hence, computational strategies such as adaptive physics refinement, where methods at one length scale spawn simulations at finer scales as needed, are required if we are to have a reliable, multiscale understanding of materials, particularly in the absence of adequate materials data that span the range of initial structures and response regimes. Such application is, and will with increasing fidelity continue to be, required in the support of the nuclear weapons and nuclear energy



An illustration of the multiscale nature of materials modeling.

programs. LANL must continue to refine the theory behind such model application.

The overarching IS&T grand challenge in materials modeling is therefore to develop a set of tools to predict, and eventually to design, the response of a material to an external stimulus with a prescribed level of accuracy. The disparate length scales that must be considered to achieve a complete understanding of materials behavior range from the atomistic characteristics (sub-nanometer detail) to the device application (typically millimeter to meter scale). There are also time scales associated with both the intrinsic length scale and phenomenon being studied, and with the particular application response of interest. Methods for quantifying the linkages between models and for properly incorporating statistical variations in the material's structures need to become inherently linked with the description of the material response. The latter is especially significant where the limits of a material's behavior (e.g., damage, failure) are likely defined by extremes in the distribution, rather than the average behavior. Success requires not only bridging these multiple length and time scales, but also seamlessly and simultaneously integrating fundamental theory, multiscale models, and predictive simulation in a manner that can be validated by experiment. This co-design vision is intended to replace the "make and observe" approach, significantly enhancing it with greater utilization of a broader range of experimental observations, and design of specific material



Polycrystal microstructure for hexagonal to orthorhombic transformation in lead orthovanadate. Variant microstructure on top; grain orientation center. Phase contrast micrograph shown at bottom.

structures that would test the limits of the models.

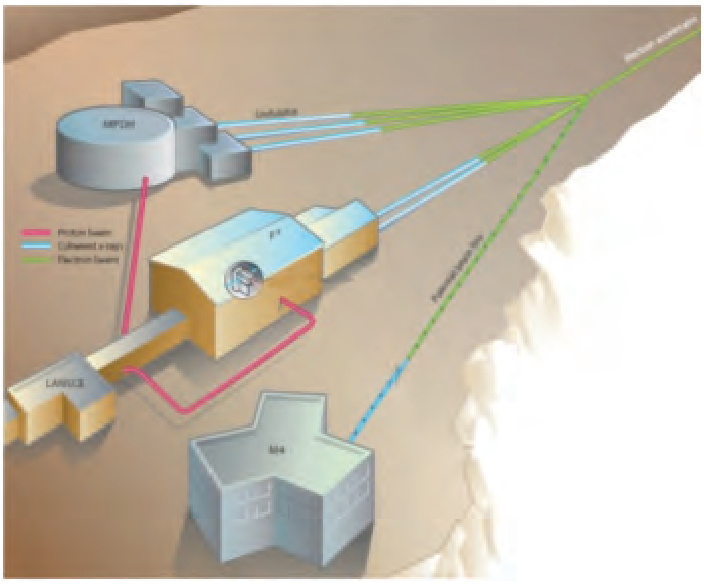
This IS&T strategy is entirely consistent with LANL's "Predicting Materials Performance" pillar, and the two pillars are mutually supportive.

LANL's Materials Strategy pillar identifies the following three scientific thrust areas that are particularly relevant to our capability to meet long-term mission needs:

- **Defects and Interfaces** – Mechanistic understanding and control of imperfections, across all appropriate length and time scales that govern materials functionality.
- **Extreme Environments** – Use of extreme environments and novel diagnostics for the characterization of materials with the ultimate goal of predicting and controlling their behavior.
- **Emergent Phenomena** – The science required to discover and understand complex and collective forms of matter that exhibit novel functionality and respond in new ways to environmental conditions, enabling the creation of materials with specified and advanced functionality.

Embedded in LANL's Materials Strategy is the proposed development of advanced experimental capabilities that target the areas of greatest uncertainty in our understanding. In particular, at length scales significantly larger than a micron, there are a number of continuum models and homogenous measurement techniques to characterize the dynamic performance of materials, although the models generally have a large empirical basis. At shorter length scales (< ~10 nm), the static and short time-scale properties of materials can be characterized or modeled with atomic precision (e.g., using molecular dynamics simulations) and there are stronger ties to first principles approaches.

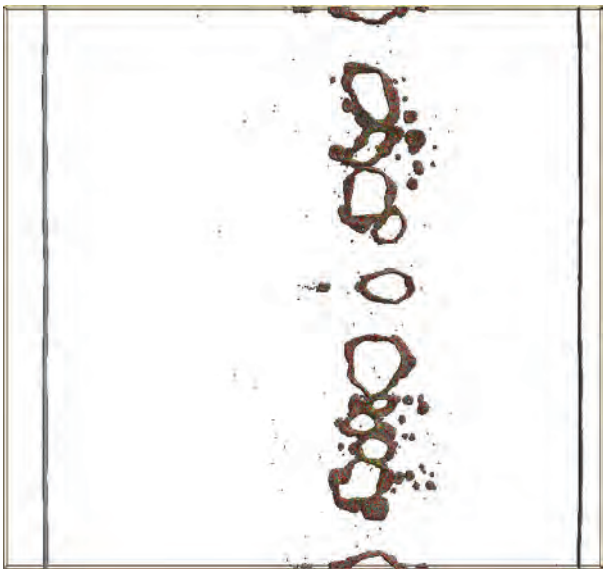
However, it is the intermediate mesoscale (~1 micron) where we have the weakest ability to characterize the dynamic response of a material. This is also where the collective behavior and self-organization of defect structures controls materials properties and performance, and where models are forced to rely on empirical extrapolations or assumptions. LANL's proposed signature facility, MaRIE, is being developed for achieving transformational materials performance in extreme environments by providing unprecedented experimental tools for



MaRIE – Schematic of the proposed Making, Measuring, and Modeling Materials Facility.

in situ, dynamic measurements of real materials in relevant environments. This data would enable a much stronger connection between the mesoscale and our current first principles models and understanding. If MaRIE is to enable a transition from an era of observation and validation to one of prediction and control, the results of these advanced observational tools must be coupled with predictive theory and a broader IS&T capability to achieve the directed synthesis of new, high-performance materials.

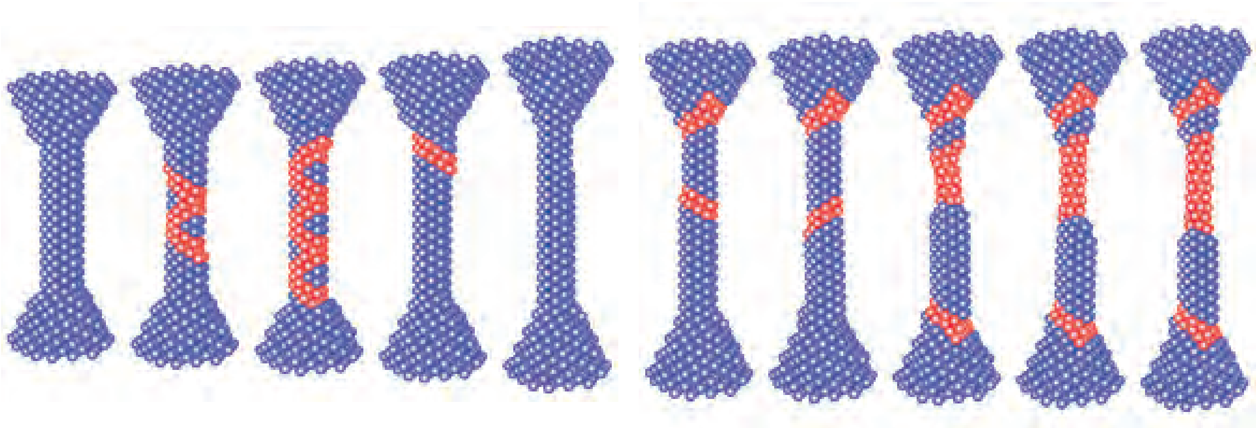
Achieving confidence in a paradigm that spans multiple decades of length and time, and is relevant to the materials science thrust areas and MaRIE, requires not only controlled mapping between scales sampled by various techniques, but also relating these to the variety of experimental measurements and making optimal use of that data and its associated uncertainties. The application of advanced IS&T technologies will have a significant impact on developing robust tools and capabilities not only for materials science but for other complex systems (e.g., biological, environmental, social, etc.) as well.



Incipient spall failure in a copper bicrystal (shocked left-to-right) by homogeneous void nucleation along the (vertical) plane of maximum tensile stress. Only undercoordinated surface atoms are shown, so that one sees the free surfaces of the impactor and target and the voids which have begun to nucleate, grow, and coalesce.

Two key questions facing materials science are: "Can we confidently describe the intermediate scales where the approximations possible for restricted time and/or spatial scales no longer apply?" and "Can we develop theories that bridge the electronic/atomistic and continuum scales?" To address these questions, we must integrate the methods of material property simulation and prediction with those of IS&T in order to maximize our understanding of materials science. This includes aspects of motivating, designing, and analyzing experiments, achieving fully integrated multiscale modeling, data fusion of experimental and modeling results (e.g., use of combinatorial and machine learning methods for predicting candidate compounds with desired properties from established data bases), techniques for dynamically analyzing massive data sets to reconstruct microstructural information, use of external statistics for rare-event prediction, and quantified measures of uncertainty applied to data extracted from both experiments and models. We will assess these methods in the context of the data extracted by the proposed MaRIE facility, and advances in our modeling capabilities in about 5 to 10 years to describe the challenge posed by the mesoscale within multiscale strategies.

The quantitative connection between atomistic simulations and continuum response models that are thus developed will have a significant impact on a broad variety of Laboratory programs. The general capability will be key to materials science problems such as constitutive properties for metals, polymers, and explosives in nuclear weapon systems, radiation damage in fuel rods, and material durability for clean energy solutions, which are likely to be among LANL's core missions in the coming decade. This capability will be critical for integrating the experimental results of the developing MaRIE program into a true materials understanding.



Early to late stages of a ParRep simulation of the stretching of a silver nanowire on Roadrunner at a temperature of 300K and a retraction velocity of 10-5 m/s. From left to right, t = 0 μs, t = 30 μs, t = 60 μs, t = 90 μs, t = 150 μs, t =165 μs, t = 180 μs, t = 195 μs, t = 210 μs, t = 225 μs. Defective sections of the wire are shown in red.

The coupling of materials science and IS&T would set LANL apart from competitors, and will demonstrate an integrated, innovative program.

Goals for Predicting Materials Performance are as follows:

- Develop capability to achieve optimal use of data and design of experiments for the enhanced quantitative prediction of materials performance.
- Establish the technical underpinnings of data-driven analysis in, for example, reconstructing microstructure from diffraction data and comparing to results from simulations, and quantified predictive capability for materials performance.
- Enable the vision of co-design through the effective integration of fundamental theory, multiscale computation, and experiment.
- Exploit the emergence of exascale-class computing by pursuing co-design strategies that integrate physics models, programming platforms, and hardware for bridging scales using adaptive physics refinement.
- Pursue applications of Complex Networks Science to the description and model solutions for the interaction of disordered structures within materials.

In addition to advancing the frontiers of materials modeling and condensed matter theory (as covered in the Materials for the Future pillar), and experiment design, data analysis, and computational co-design, our strategy is to integrate these individual components to better predict materials behavior. We will build the necessary collaborations and teams to carry out this integration.

Application Area: Situational Awareness

Situational Awareness is the perception of elements within a potentially large domain of time and space, the comprehension of their meaning, and the projection of their status in the near future. It concerns perception of the environment critical to decision makers in complex, dynamic areas such as nonproliferation, counterterrorism, military command and control, space, cyber infrastructure, and emergency response.

The nation's security increasingly depends on acquiring and processing information as we attempt to recognize and respond to complex threats. These threats often have ephemeral and subtle signatures that arise at unpredictable times and locations. While the sensors being built to address national needs are diverse, the common threads are exponentially growing data rates and the need for spatially distributed, persistent monitoring. To effectively mitigate a threat, that threat must first be recognized and interrogated as it emerges and evolves. However, human analysts do not have the memory, patience, stamina, or reaction time required to monitor huge data flows, recognize important and subtle variations, and promptly respond with follow-up observations and actions. To be effective, next-generation systems will need to be capable of recognizing rare events and anomalies in a torrent of data. They must also respond in near real-time by formulating queries and priorities, making the relevant follow-up observations, and folding the new information back into the process to determine the next action. They must be autonomous systems that incorporate scalable, distributed, and intelligent information science. Moreover, these systems must provide quantitative assessments of the quality of the predictions and recommendations they make.

As a JASON report on "Data Analysis Needs" (2008) stated: "DoD/IC data requirements are certainly significant, but not unmanageable given the capabilities of current and projected storage technology. The key challenge will be to adequately empower the analyst by matching analysis needs to data delivery modalities." Solving this problem requires a program of R&D focused on real data and real users.

The fundamental challenge for information science is to expose, extract, and predict phenomena of interest from an increasingly diverse set of sensors and information sources and from an increasingly cluttered and unpredictable background. To address this challenge we will require computational solutions that draw from a diverse set of capabilities including sensor modeling, modality-specific data modeling, analytics modeling, knowledge modeling, and large-scale data management, as well as visualization, human computer interaction, and decision support sciences.

Specific domains of Situational Awareness include the following:

- **Situational Awareness for Proliferation Threats** – Threats related to proliferation include the production, transport, or use of weapons of mass destruction, particularly as related to nuclear materials and activities. Situational awareness for this area seeks to perform monitoring and surveillance to detect and characterize nuclear-related activities as early as possible. This will include the monitoring of materials, facilities, and persons related to these proliferation activities and monitoring developing scientific and technical capabilities related to proliferation by mining data from the LANL Research Library and the ODNI (Office of the Director of National Intelligence) Open Source Center.

RAPTOR, Thinking Telescope: Fast response, persistent monitoring

- **Wide Area Surveillance Science** – The primary objective of LANL's Wide-Area Surveillance Science (WASS) is to mitigate terrorist influence through the use of technologies and analytical tools integrated with wide-area imagery. Examples include IED emplacement detection, forensic analysis of terrorist/insurgent events, and operational support of US personnel in-theater. LANL's advancement in this area will also contribute to homeland security, disaster response, infrastructure monitoring, and resource management.

- **Space Situational Awareness** – Space Situational Awareness (SSA) is the ability to determine what is happening in space that can affect our space systems on a "24/7" basis. This includes items such as 1) where all objects in space are, 2) what they are doing, 3) what the space and ionospheric environments are doing and how they can affect our space operations, 4) what is the status of the ground and information infrastructures that support space operations, and 5) identification of any man-made or natural threats to our space systems. SSA is essential to provide the timely knowledge for mitigation and policy options to protect the services we derive from space systems, the infrastructure for which we have often taken for granted. Timely hurricane forecasting, satellite TV and radio, GPS navigation, views from space on Google Earth, tracking our FedEx and UPS packages, timely banking transactions, satellite and cell phone service, emergency beacon location, missile warning, and nuclear treaty monitoring are just some of the diverse ways that our commercial, civilian, and military space infrastructures affect our lives.

- **Cyber-situational Awareness** – Our national cyber infrastructure is under relentless attack from a variety of adversaries. Unless we can guarantee robust responses to these attacks, the nation will continue to be more and more vulnerable to a cyber version of Pearl Harbor. Such an attack would have devastating effects, not limited to the severe economic damage that we have already witnessed with recent, well-known worms and viruses. To address this serious cyber challenge, LANL is developing the scientific underpinnings to cyber security. This scientific approach is based on measurement, modeling, and experiment (through both simulation and actual instantiation on the LANL network). By developing objective measures of performance, and being able to model, predict, and validate performance, we will change the course of cyber security from a path of Edisonian improvement and response onto a path of more scientific objectivity. In addition to providing the foundations for "classical" cyber security science, LANL is also carrying out R&D in very promising avenues of quantum-enabled cyber security.

In addition to these four programs in Situational Awareness, LANL has and must maintain active programs in Structural Health Monitoring and Biological/Pathogen Monitoring. While these areas are distinct, many of the IS&T capabilities necessary to assess the situations are common ingredients.

Goals for Situational Awareness include the following:

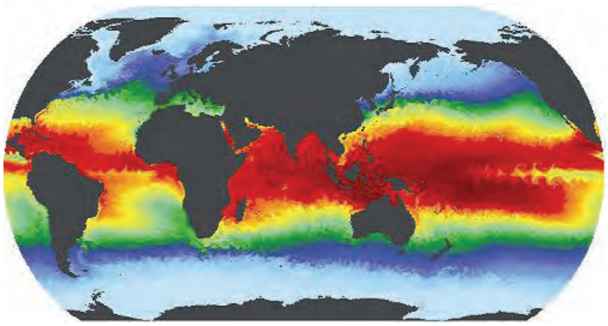
- Develop and execute a beginning-to-end pilot project for a specific global security application to demonstrate LANL capabilities in optimal inference/situational awareness.
- Provide scientific underpinnings for cyber security research via interdisciplinary collaborations and programmatic opportunities.
- Grow the cyber security component of LANL's QIS capability within the National Nuclear Security Administration (NNSA) and other government organizations.
- Maintain and further develop LANL's world-class expertise in specific technical components relevant to situational awareness, such as sensor, algorithm, and application niches.

• Provide the scientific foundation for unifying situational awareness capacities across the Laboratory.

• Focus on coupling the ability to collect large quantities of data with emerging IS&T tools to take this data to information and then actionable knowledge. We will form cross-disciplinary and cross-laboratory teams to achieve this coupling.

Application Area: Energy-Climate Impacts and Energy Infrastructure

Society's energy use and its impacts on our global and regional climate pose scientific, technological, and societal challenges. The IS&T aspects of these challenges will involve efforts in the following areas:



Temperature at 15 m depth from a high resolution ocean simulation using the Parallel Ocean Program (POP).

• Quantification of greenhouse gas (GHG) emissions and offsets for policy makers and scientists.

• UQ for global and regional climate prediction.

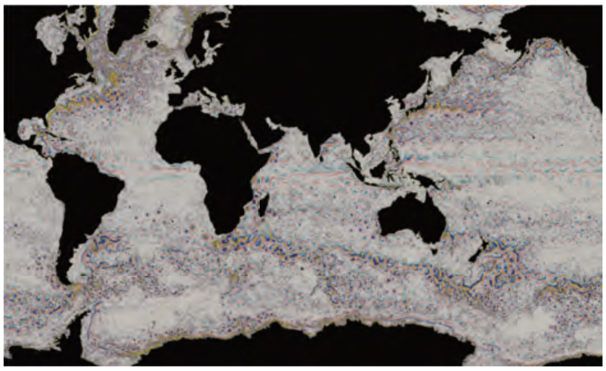
• Infrastructure modeling, design, and optimization

These IS&T efforts are entirely consistent with LANL's "Science of Signatures" (SOS) pillar and strategy, and the two pillars are mutually supportive.

• **Quantification of GHG Emissions** – There is scientific consensus that understanding and control of GHG emissions is essential to mitigating and managing impacts of climate change. International policy is focused on how rapid these emission reductions must be in order to stabilize atmospheric carbon dioxide at levels that prevent dangerous interference with our climate system. International treaties and/or cap and trade policies are imminent nationally and globally. Presently, volunteer reporting by states and nations of their annual energy consumption to the DOE and IEA (International Energy Agency) is used to compute their carbon dioxide emissions. This methodology tracks the amount of coal, oil, and gas consumed and assumes combustion efficiencies of energy systems (power, transportation) in order to derive emissions. The efficiencies of transportation systems are variable, and the amount of biofuel used is highly uncertain. Empirical rules to calculate emissions use approximations that suffer from large uncertainties. Therefore, a more rigorous system for quantifying emissions and offsets is required. Working in concert with the SOS pillar, the IS&T pillar will provide the necessary integration of Laboratory capabilities in theory, modeling, simulation, and data analysis to help build the GHG Information System (GHGIS).

• **Uncertainty Quantification for Climate** – The earth's climate system contains several modes of natural variability and response to human changes. To assess this internal variability and the related uncertainties in climate projections, climate modelers simulate several instances of each climate change scenario (choice of future human emissions). These differences in simulation between members of each ensemble and comparisons with similar studies done by each of the many International Panel on Climate Change modeling groups provides an estimate of the internal variability due to both natural modes of variability and the sensitivity to model parameterizations and assumptions.

Global climate models are computationally intensive, requiring months of computer time on high-end computing resources. Regional



Vorticity from a high resolution simulation that resolves ocean eddy structure.

climate models require great care and quality control in analyzing the relative biases that are either internal to the model or forced by the global climate data at the lateral boundaries. It is likely in the near term that most global and regional climate data in a decision support system will come from existing climate data sets from previously computed ensembles of future climate projections. Methods for combining ensemble data to reduce overall uncertainty will be required. In addition, it may be possible to use lower-order models to capture some aspects of climate sensitivity without resorting to full dynamical models. For example, the globally averaged surface temperature and some patterns in the surface temperature response are well correlated to greenhouse gas concentrations and might be represented in a simpler model. Precipitation, on the other hand, shows a more complex and more highly localized response.

The IS&T pillar will work to develop more rigorous and formal uncertainty quantification, parameter estimation, and error propagation analyses which are required throughout the climate and regional modeling community. The global climate community has made some use of ensembles of simulations to quantify uncertainty, but more stringent statistical methods need to be introduced for more rigorous uncertainty measures. In regional climate modeling, ensembles are also being used and qualitative estimates of the level of uncertainty have begun, largely in assessing the amount of bias introduced by boundary conditions versus bias introduced by the regional models themselves. This work must continue, but with more quantitative measures derived through formal statistical methods.

- **Infrastructure Modeling, Design, and Optimization** – The basic structure of the US electrical power grid has remained unchanged for 100 years. It is a hierarchical, centrally controlled structure that assumes that power is generated solely from large central facilities, that power is abundant, and that power generation is relatively benign. Stability is attained through redundancies and a highly controllable generation that reacts to problems such as demand fluctuations and outages, rather than anticipating and avoiding them.

A number of significant developments have fundamentally challenged the assumptions under which the current grid was designed. Demand has risen sharply, and will continue to do so. Scientific analysis has shown that we need to curtail or eliminate our use of the most abundant sources of power (fossil fuels), in order to avoid severe environmental consequences. Alternative, renewable sources of energy such as wind and solar are available, but are often distributed and intermittent, and thus difficult to control.

These developments are forcing a paradigm shift in the structure of the future grid. The essence of this shift is that tools of information technology are being introduced to improve the efficiency and stability of the grid, and to facilitate the use of distributed generation from renewable energy sources. The future grid, in which modern sensors, communication links, and computational power are used in concert to improve grid efficiency, stability, and flexibility, has become known as the "smart grid." The administration and the DOE both recognize the need for a smart grid, and the total national investment in the grid is expected to exceed $2 trillion over the next 20 years.

Much of the hardware that will enable this revolution is in development or already exists—"smart" meters and appliances that respond to pricing signals, distributed wireless sensor networks, improved batteries for plug-in hybrids that enable distributed storage, and so forth. What is largely lacking, and what this application area intends to provide, is a fundamental understanding of how best to wed IS&T to the electrical grid. The smart grid is now at the stage of telephony when the first telephones were invented – the potential was obvious, but the realization of this potential required extensive development in communication and switching theory, of which information theory itself was a spinoff. In the same way, information technology will be essential to realizing the full potential of the smart grid.

We will develop algorithms for optimal design of the nation's next generation power transmission and distribution system that will incorporate the new realities of the grid. Our ultimate goal is an innovative suite of real-time capabilities for detecting and preventing instabilities and outages and accommodating intermittencies in production, and a state-of-the-art framework for analyzing and designing the smart grid and associated control network. Success requires careful and deliberate integration among the three capability thrusts in the IS&T pillar – data intensive scientific computing, computational co-design, and complex networks.

Goals for Energy Climate Impacts and Energy Infrastructure are as follows:

- Develop scientific underpinnings for IS&T aspects of smart grid.
- Work with climate scientists to develop advanced UQ capability for climate prediction.
- Foster development of cross-disciplinary data-fusion capability, including climate-energy-geospatial systems.
- Participate in the emerging discipline of energy impacts science through the Energy Climate Impacts program development effort.
- Develop new methods for quantifying the uncertainties in GHGs for the program development effort tasked with developing a new GHGIS.

Our strategy for achieving these goals will focus on integrating our considerable expertise in inference and uncertainty quantification, computational science, and climate and environmental science. We are building the necessary collaborations and teams to carry out this integration.

Summary

IS&T has become a cornerstone of science in the 21st century. High-resolution experiments, wide-area observations, distributed sensor networks, and large-scale computational resources generate petabytes of data on a daily basis, and the analysis of these massive datasets has, in some fields, led to a shift in how science is done. LANL is committed to defining and exploiting the full potential of this new paradigm to solve problems of national importance by creating a strong and coherent effort in IS&T.

Glossary

Term	Meaning
CNLS	Center for Non-linear Science
DoD	Department of Defense
DOE	Department of Energy
ECI	Energy and Climate Impacts
GHG	Greenhouse Gas
GHGIS	Greenhouse Gas Information System
IC	Intelligence Community
IEA	International Energy Agency
IED	Improvised Explosive Devices
IS&T	Information Science and Technology
LANL	Los Alamos National Laboratory
LDRD	Laboratory Directed Research and Development
LSST	Large Synoptic Survey Telescope
MaRIE	Matter-Radiation Interactions in Extremes
NNSA	National Nuclear Security Administration
NPR	Nuclear Posture Review
ODNI	Office of the Director of National Intelligence
QIS	Quantum Information Science
R&D	Research and Development
SOS	Science of Signatures
SSA	Space Situational Awareness
ST&E	Science, Technology, and Engineering
UQ	Uncertainty Quantification
UQV&V	Uncertainty Quantification Verification and Validation
V&V	Verification and Validation
WASS	Wide-Area Surveillance Science

LA-UR 11-00115